

Détermination du taux d'huile des graines de cotonnier par spectrophotométrie dans le proche infrarouge

C. Marquié, A.-M. Tessier

CIRAD-CA, laboratoire de technologie cotonnière, BP 5035, 34032 Montpellier Cedex 01, France.

Résumé

Les auteurs donnent les résultats d'une étude préliminaire sur l'étalonnage d'un spectrophotomètre dans le proche infrarouge pour déterminer le taux d'huile des graines de cotonnier. La validation de l'étalonnage met en évidence une corrélation insuffisante entre les valeurs de référence et celles obtenues par spectrophotométrie (coefficient de détermination : 0,79). Les auteurs proposent des améliorations à apporter dans le conditionnement et la mesure des échantillons.

MOTS-CLES : graines de cotonnier, huile, spectrophotométrie proche infrarouge.

Introduction

La spectrophotométrie proche infrarouge (SPIR) est une technique de dosage et de caractérisation moléculaire dont les premiers développements par le Ministère de l'agriculture des Etats-Unis remontent à plus de 25 ans. Elle a rencontré un fort intérêt dans l'industrie céréalière pour le dosage de l'amidon en particulier (WILLIAMS et NORRIS, 1990).

Nous avons employé la SPIR dans le but de déterminer la teneur en huile d'échantillons de graines de cotonniers provenant d'essais variétaux, obtenus dans le cadre de programmes d'amélioration variétale.

Actuellement, la plupart des méthodes de détermination du taux d'huile d'un échantillon repose sur l'extraction des lipides par un solvant organique apolaire tel que

l'hexane. Ces méthodes, assez fastidieuses, impliquent la réalisation de nombreuses pesées et une durée d'extraction relativement longue (une vingtaine d'heures, dans le cas de la méthode « russe » publiée par BOURÉLY, en 1982).

Par rapport à ces méthodes, les avantages de la SPIR sont nombreux. La SPIR ne nécessite pas de compétence particulière pour l'utilisateur une fois l'étalonnage réalisé. Pour des constituants qui présentent des caractéristiques spectrales suffisamment spécifiques (eau, lipides, glucides), la SPIR est intéressante en raison de la rapidité de la réponse, de la faible quantité de produit (récupérable) nécessaire, du coût de revient des analyses et de la fiabilité des appareils conçus pour fonctionner souvent dans un environnement industriel.

Principe de la SPIR

Le principe de la SPIR est basé sur l'action des radiations électromagnétiques au niveau des molécules organiques.

En effet, les électrons de la liaison covalente (de type σ ou π) absorbent l'énergie de la radiation proche infrarouge quand la fréquence de celle-ci est identique à celle de la

vibration pour passer à un niveau énergétique supérieur. L'énergie absorbée dépend des atomes constituant la liaison et de leur environnement dans la molécule. L'excitation des électrons a pour effet d'augmenter l'amplitude des vibrations et de modifier l'angle formé par les deux liaisons. Les vibrations les plus significatives sont celles des liaisons constituées d'atomes de masses atomiques très différentes

(C-H, C-O, N-H, C=O, C=N...). Chaque type de vibration est produit par une radiation de fréquence ou de longueur d'onde spécifique.

Chaque constituant d'un produit organique présente donc un spectre d'absorption caractéristique dans le domaine infrarouge, avec des bandes spécifiques.

Principe de la mesure

La lumière incidente (émise par une source lumineuse, fig. 1) est sélectivement absorbée ou réfléchi : une partie est absorbée par les liaisons des constituants, l'autre partie est réfléchi de deux façons différentes.

La réflexion spéculaire correspond à la lumière réfléchi directement par la surface des particules constituant l'échantillon. Cette réflexion dépend entre autres de la taille, de l'état de surface et de la forme des particules.

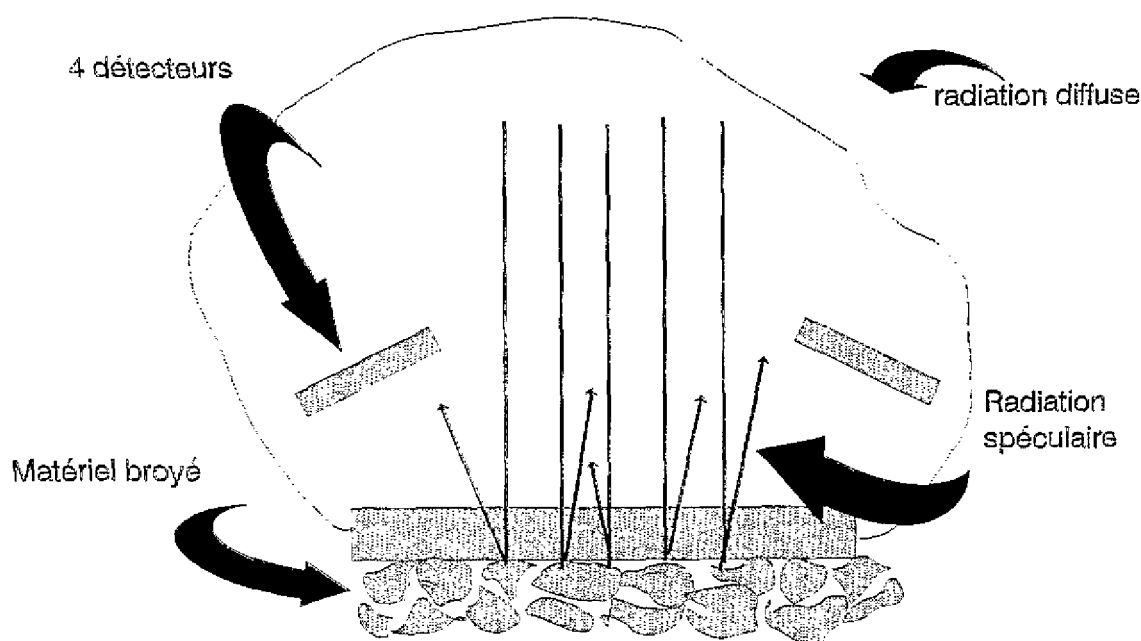


Figure 1

Principe de l'analyse par réflexion dans le proche infrarouge. L'information chimique est contenue uniquement dans la radiation diffuse.

Principle of analysis by near infrared reflection. The chemical information is contained solely in the diffuse radiation.

La réflexion diffuse correspond à la lumière non absorbée qui, après un trajet aléatoire dans l'échantillon, est renvoyée à la surface de l'échantillon et réfléchi. Son intensité dépend essentiellement du type et du nombre de liaisons chimiques présentes dans le produit analysé.

Les appareils d'analyse par réflectance dans le proche infrarouge mesurent la totalité de la quantité d'énergie réfléchi par un échantillon soumis à une radiation lumineuse dont le domaine spectral est compris entre 1100 et 2500 nm. Le rapport entre lumière émise et lumière réfléchi est fonction de la nature des composés présents dans le produit. Le spectre d'absorbance représente la somme des spectres de nombreux constituants et comporte souvent des bandes qui se recouvrent. L'exploitation des spectres nécessite alors, pour chaque constituant à analyser, de tenir compte de la réflexion à plusieurs longueurs d'onde. Ce qui permet de pallier les interférences, c'est-à-

dire l'absorbance des autres constituants (DAVIES et GRANT, 1987 ; WETZEL, 1983).

Pour l'étalonnage de l'appareil, la réflectance du rayonnement est mesurée pour de nombreux échantillons de référence, à différentes longueurs d'onde.

A l'aide des valeurs d'absorbance obtenues et des valeurs de référence données par une autre méthode de laboratoire, une analyse de régression linéaire multiple par la méthode des moindres carrés est réalisée, pour déterminer une équation de prédiction. Celle-ci peut être appliquée pour calculer le taux d'huile d'échantillons inconnus.

Pour la validation de la méthode, d'autres échantillons de graines, indépendants de la gamme ayant servi à l'étalonnage et dont la teneur en lipides a été préalablement déterminée, sont analysés par la SPIR. Les valeurs prédites sont comparées aux valeurs de référence.

Matériels et méthodes

Matériels

Les mesures de spectres dans le proche infrarouge sont effectuées avec un appareil Nirsystème 4500 Perstorp couplé à un ordinateur Intel i386 302-20 (mis à notre disposition par le laboratoire de M. Guyot, CIRAD-CP). L'ensemble du système est exploité par le logiciel Isi 205, mis au point par SHENK et WESTERHAUS (1991).

Trois cent trente et un échantillons de graines de cotonnier nues et broyées de provenances diverses (Mali, Burkina-Faso, Togo, Sénégal, Thaïlande, Bénin et Brésil) ont été utilisés pour réaliser l'étalonnage.

Par ailleurs, 82 échantillons de graines de cotonnier, préparés dans les mêmes conditions que celles employées au cours de l'étalonnage, sont utilisés pour valider la méthode.

Les mesures d'absorption dans le proche infrarouge, pour chaque échantillon de graines broyées et non conditionnées (humidité relative non contrôlée), sont réalisées par réflexion entre 1100 et 2500 nm. Le spectre de chaque échantillon est la moyenne de 32 spectres obtenus pendant la rotation de la chambre de mesure.

La détermination du taux d'huile de chaque échantillon de graine, servant à l'étalonnage et à la validation de la méthode spectrophotométrique, est réalisée par la méthode «russe» (BOURÉLY, 1982).

Méthodes d'étalonnage

L'étalonnage est réalisé à partir d'un fichier de 331 spectres, structuré par les procédures statistiques suivantes.

- Analyse en composantes principales (ACP). L'ACP est utilisée pour réduire le nombre de variables réelles, c'est-à-dire le nombre de longueurs d'onde comprises entre 1100 et 2500 nm, à un nombre de variables compatible afin que les calculs de distance de Mahalanobis ne soient pas trop lourds.

- Distance généralisée de Mahalanobis (WILLIAMS et NORRIS, 1987). Cette distance est, pour un spectre donné, sa distance en unités d'absorbance par rapport à la moyenne de tous les spectres (centre de gravité) pour une ou plusieurs longueurs d'onde données. Ces distances mesurées en terme de «H» doivent rester inférieures à 3 (WESTERHAUS et SHENK, 1991).

Dans le cadre de notre étude 5 échantillons, pour lesquels le «H» est supérieur à 3, sont considérés comme atypiques et rejetés du fichier.

Afin de déterminer les échantillons les plus représentatifs de la population qui doivent être utilisés pour effectuer l'étalonnage, une sélection statistique est réalisée par un calcul de distances de Mahalanobis : ceci pour identifier et sélectionner les spectres qui ont le plus de voisins les plus proches.

Le traitement mathématique a permis de sélectionner, pour l'étalonnage, 244 échantillons qui sont dosés par ailleurs par la méthode de référence.

L'étalonnage permet d'associer une valeur mathématique à chaque spectre, qui sera elle-même associée à la valeur de référence.

Résultats et discussion

Détermination de l'équation d'étalonnage

Le logiciel Isi 205 offre la possibilité de choisir parmi plusieurs paramètres qui influent sur les performances de l'étalonnage (choix du traitement mathématique, de la méthode de calcul de la régression, du domaine de longueur d'onde utilisé, du nombre de validations croisées ...). Les paramètres sélectionnés qui nous ont permis d'obtenir les meilleurs résultats sont les suivants :

- étendue du spectre, 1100 nm - 2500 nm ;
- traitement mathématique :
 - dérivée seconde,
 - pas, 5 nm,
 - lissages, 5 et 1,
- méthode de régression MPLS («modified partial least square») ;
- validation croisée, 4 ;

- nombre de termes de l'équation de prédiction, 12.

La méthode MPLS permet de réduire les erreurs dues au calcul matriciel des distances de Mahalanobis.

Les variations de spectres dues aux effets de surface, intervenant lorsque le signal de réflectance est directement utilisé, peuvent être réduites en utilisant les dérivées premières et secondes de Log de (1/R) (COOPER, 1983 ; NORRIS et WILLIAMS, 1984). L'utilisation des dérivées secondes donne généralement de meilleurs résultats (DARDENNE et BISTON, 1986).

Si le broyage des produits dans des conditions rigoureuses de granulométrie est un facteur prépondérant (OSBORNE, 1983), l'utilisation de dérivées secondes permet d'estomper largement les différences.

Le spectre de chaque échantillon est enregistré avec un pas de 5 nm. Pour avoir la meilleure précision, le pas doit être le plus petit possible (fig. 2).

Les lissages successifs permettent de diminuer le bruit de fond.

La procédure d'étalonnage consiste à choisir trois quarts des échantillons sur lesquels un modèle de régression multiple est développé (MPLS). Celui-ci est ensuite appliqué au quart des échantillons restants. La procédure est répétée trois fois, en réservant toujours un quart des

échantillons pour la validation. Les écarts des quatre prédictions sont cumulés pour obtenir une erreur standard de validation, dont le minimum fixe le nombre de facteurs à introduire dans le modèle. Le modèle final est alors recalculé sur l'ensemble des échantillons.

Le nombre maximal de termes de l'équation est fixé à 12 par l'expérimentateur. Lors de la validation croisée, le logiciel choisit le nombre de termes nécessaires permettant d'établir la meilleure équation d'étalonnage. Une bonne évaluation du nombre de termes nécessaires est de un terme pour 10 échantillons.

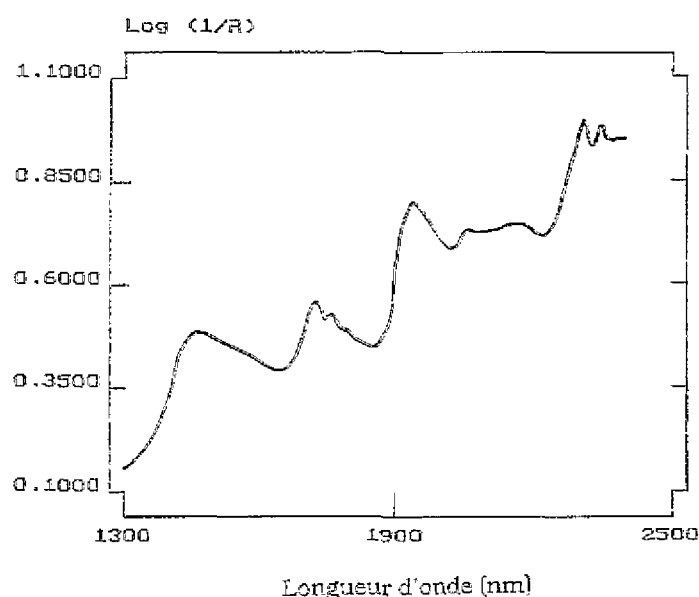


Figure 2

Spectre d'absorption dans le proche infrarouge d'un échantillon de graines de cotonnier (R étant la réflectance).
Absorption spectrum in the near infrared of a sample of cotton seeds (R , reflectance).

La validation croisée élimine encore un certain nombre de spectres dont l'écart entre les valeurs prédites et les valeurs de référence est supérieur à 2 % d'huile (valeur de «T», fixée par l'expérimentateur). Dix échantillons présentant un «H» ou un «T» trop important sont rejetés.

Les résultats de l'étalonnage sont donnés dans le tableau I.

Validation

L'étalonnage est validé en analysant 32 échantillons indépendants.

Le coefficient de corrélation entre les valeurs de référence et les valeurs prédites est de 0,888, soit un coefficient de détermination de 78,9 %.

La validation de cet étalonnage montre que la corrélation entre les valeurs de référence et les valeurs prédites

n'est pas mauvaise, mais insuffisante pour utiliser cette équation de prédiction.

Propositions d'amélioration de la méthode

Du fait que les paramètres des équations sont déterminés à partir de l'ensemble du spectre de chaque échantillon, la variation de la teneur en eau peut influencer sur l'analyse spectrophotométrique.

Nous suggérons de procéder à un nouvel étalonnage de la détermination de la teneur en huile à partir de 120 échantillons de graines de cotonnier issus d'une banque de génotypes, conditionnés à 0 ou à 80 % d'humidité relative.

Le broyage des graines de cotonnier sera standardisé : environ 20 g de graines seront broyées au moulin à café en 6 périodes de 5 secondes, entre lesquelles la matière sera homogénéisée au moyen d'une spatule.

TABLEAU 1

Caractéristiques de l'étalonnage réalisé sur 234 échantillons de graines de cotonnier délintées.

Characteristics of the calibration carried out on 234 delinted cotton seed samples.

Traitement mathématique	2-5-5-1
Méthode de régression	MPLS
Cross-validation	4
Nombre de termes	12
N	234
MOY	25,28
SEC	0,53
RSQ	0,98
SECV	0,65
1-VR	0,93
BIAS	0,32
SEP	0,61

N : nombre d'échantillons sélectionnés pour l'étalonnage ;

MOY : moyenne de la teneur en huile de la population ;

SEC : erreur standard de calibration ;

RSQ : coefficient de détermination de calibration, obtenu d'après la droite établie entre les valeurs de référence et les valeurs SPIR ;

SECV : erreur standard de cross-validation. Elle doit théoriquement tendre vers 0. Cependant, d'après la littérature cette valeur sera dans le meilleur des cas comprise entre 0,5 et 0,6 ;

VR : variance ;

BIAS : moyenne des différences entre les valeurs obtenues par la SPIR et les valeurs de référence pour plusieurs échantillons appelés lots de validation ou de vérification ;

SEP : erreur standard de prédiction ou de performance.

N: number of samples selected for calibration

MEAN: mean oil content of the population

CSD: calibration standard deviation

RSQ: calibration coefficient of determination obtained from the straight line drawn up between the reference values and the NIRS values.

CCSD: cross-checking standard deviation. In theory, this should tend towards 0. However, according to the literature, this value will be between 0.5 and 0.6, at best.

VR: variance

BIAS: mean of the differences between the values obtained by the NIRS and the reference values for several samples called validation or verification batches.

PSD: prediction or performance standard deviation.

La détermination de la teneur en huile par la méthode de référence (méthode « russe ») sera obtenue à partir de 8 analyses, dont on considèrera la moyenne des résultats pour avoir une erreur relative de \pm ou $-0,25$ % d'huile. La valeur de « T » (différence tolérée entre les valeurs obtenues et les valeurs prédites) sera choisie en fonction de l'erreur relative observée sur les résultats obtenus à partir de la méthode de référence.

La validation de l'étalonnage sera faite sur 40 échan-

tillons n'ayant pas servi à l'étalonnage et la corrélation entre les valeurs NTR et les valeurs de référence sera déterminée. Si l'étalonnage ne s'avère pas satisfaisant, les données spectrales de ces 40 échantillons seront rassemblées au fichier initial et un nouvel étalonnage sera réalisé. Une validation devra alors être effectuée avec 40 autres échantillons etc.

Il est probable que le nombre d'échantillons nécessaires pour obtenir un meilleur étalonnage soit supérieur à 120.

Conclusion

Les résultats de cette étude préliminaire montrent que la détermination du taux d'huile des graines de cotonnier par la SPIR est réalisable.

Toutefois, bien que les caractéristiques d'étalonnage obtenues par la « cross validation » soient correctes (RSQ > 0,9), la validation de la méthode met en évidence un niveau

de corrélation entre les valeurs de référence et les valeurs SPIR trop faible pour les exigences du laboratoire.

Pour réaliser l'étalonnage et ensuite les analyses, il convient de mieux standardiser le mode de préparation et de conditionnement des échantillons dont les caractéristiques physiques influent sur les analyses NTR.

Notre laboratoire conduit actuellement des travaux pour améliorer cette méthode spectrophotométrique et pourra, dès la fin de l'année 1994, l'utiliser régulièrement pour la détermination du taux d'huile des essais variétaux.

La spectrophotométrie dans le proche infrarouge présente un domaine d'utilisation très large pour le dosage des composants organiques. Nous envisageons, d'ores et déjà, de l'employer pour déterminer la teneur en protéines et en gossypol des graines de cotonnier.

Références bibliographiques

- BOURELY J., 1982.- Observations sur le dosage de l'huile des graines de cotonnier. *Coton Fibres Trop.*, 37, 2, 185-196.
- COOPER P.J., 1983.- NIR analysis for process control. *Cereal Foods World*, 28, 241-245.
- DARDENNE P., BISTON R., 1986.- L'analyse discriminante appliquée au contrôle granulométrique des produits soumis à la spectrométrie dans le proche infrarouge. *Sci. Aliments*, 6, 1, 1-14.
- DAVIES A.M.C., GRANT A., 1987.- Review : Near infra-red analysis of food. *Internat. J. Food Sci. and Technol.*, 22, 3, 191-207.
- KOP J., BONNET M., RENOU J.P., 1986.- Contrôle non destructif. Application des techniques utilisant l'infrarouge et la RMN au contrôle de la viande et des produits carnés. *Spectra*, 14, 110, 37-43.
- NORRIS K.H., WILLIAMS P.C., 1984.- Optimization of mathematical treatments of raw near infrared signal in the measurement of protein in hard red spring wheat. I : Influence of particle size. *Cereal Chem.*, 61, 2, 158-165.
- OSBORNE B.G., 1981.- Principles and practice of near infra-red (NIR) reflectance analysis. *J. Food Technol.*, 16, 1, 13-19.
- OSBORNE B.G., 1983.- Investigation of the performance of an improved calibration for the determination of protein in UK home-grown wheat by near infrared reflectance analysis. *J. Sci. Food Agric.*, 34, 12, 1441-1443.
- PANFORD J.A., DEMAN J.M., 1990.- Determination of oil content of seeds by NIR: influence of fatty acid composition on wavelength selection. *J.A.O.C.S.*, 67, 3, 473-482.
- WESTERHAUS M.O., SHENK J.S., 1991.- Near infrared software. NIRS 3. *Infrasoft International, Nirsystems*. Silver spring.
- WEIZELD L., 1983.- Near infrared reflectance analysis : sleeper among spectroscopic techniques. *J. Anal. Chem.*, 55, 12, 1165-1176.
- WILLIAMS P.C., NORRIS K.H., 1987.- Near infrared technology in the agricultural and food industries. *Amer. Ass. Cereal Chem.*, USA, 330 p.
- WILLIAMS P.C., NORRIS K.H., 1990.- Near infrared technology in the agricultural and food industries. *Amer. Assoc. Cereal Chem.*, 35-55.

Determination of cotton seed oil content using near infrared spectrophotometry

C. Marquié, A.-M. Tessier

Abstract

The authors give the results obtained in a preliminary study on the calibration of a near infrared spectrophotometer to determine the oil content of cotton seeds. Validation of the calibration

revealed an insufficient correlation between the reference values and the spectrophotometric values obtained (coefficient of determination: 0.79). The authors propose improvements to be made in sample preparation and measurement.

KEYWORDS: cotton seeds, oil, near infrared spectrophotometry.

Introduction

Near infrared spectrophotometry (NIRS) is a quantification and molecular characterization technique first developed by the U.S. Ministry of Agriculture more than 25 years ago, and which aroused great interest in the cereals industry for starch quantification, for example (WILLIAMS and NORRIS, 1990).

We used NIRS with a view to determining the oil content of cotton seed samples from varietal trials, obtained under varietal improvement programmes.

Most of the methods currently used to determine the oil content of a sample are based on lipid extraction by an

apolar organic solvent, such as hexane. These methods are fairly laborious, involving numerous weighing operations and a relatively long extraction time (around 20 hours in the case of the «Russian» method published by BOURELY in 1982).

Compared to these methods, NIRS offers many advantages. It does not require any specific skills once calibration has been carried out. For constituents with sufficiently specific spectral characteristics (water, lipids, carbohydrates), the interest of NIRS lies in its rapid response, the use of only small quantities of product (recoverable), a low cost price and reliable equipment designed for intensive use in an industrial environment.

The NIRS principle

NIRS is based on the action of electromagnetic radiation on organic molecules.

In fact, the electrons of the covalent bond (type σ or π) absorb near infrared radiation when its frequency is identical to that of the vibration to pass to a higher energy level. The absorbed energy depends on the atoms making up the bond and their environment within the molecule. Excitation of the electrons increases the amplitude of the vibrations and

modifies the angle formed by the two bonds. The most significant vibrations are those of the bonds making up atoms with very different atomic weights (C-H, C-O, N-H, C=O, C=N...). Each type of vibration is produced by radiation with a specific frequency or wavelength.

Each constituent of an organic product therefore has a characteristic absorption spectrum in the infrared range with specific bands.

Measuring principle

Incident light (emitted by a light source, fig. 1) is selectively absorbed or reflected; one part is absorbed by the bonds of the constituents, the other part is reflected in two different ways.

Specular reflection corresponds to light reflected directly by the surface of the particles making up the sample. Among other things, this reflection depends on the size and condition of the surface and the shape of the particles.

Diffuse reflection corresponds to non-absorbed light which, after a random course in the sample, is sent back to the surface of the sample and reflected. Its intensity primarily depends on the type and number of chemical bonds in the product analyzed.

Near infrared reflectance analysis instruments measure the total amount of energy reflected by a sample subjected to light radiation with a spectral range of between 1,100 and 2,500 nm. The ratio between emitted light and reflected light depends on the type of compounds in the product. The absorption spectrum is the sum of the spectra of

numerous constituents and often consists of overlapping bands. To use the spectra, it is therefore necessary to take into account the reflection at several wavelengths, for each of the constituents to be analyzed. This makes it possible to moderate interference, i.e. absorption by the other constituents (DAVIES and GRANT, 1987; WETZEL, 1983).

To calibrate the equipment, radiation reflectance is measured for numerous reference samples, at different wavelengths.

Using the absorption values obtained and the reference values given by another laboratory method, a multiple linear regression analysis using the least squares method is carried out to obtain a predictive equation. This can be used to calculate the oil content of unknown samples.

In order to validate the method, other seed samples, independent of the range used for calibration and whose oil content has been determined beforehand, are analyzed by the NIRS. The predicted values are compared to the reference values.

Material and methods

Material

Near infrared spectrum measurements were carried out on Nirsystem 4500 Perstorp type apparatus linked to an Intel i386 302-20 computer (made available to us by Mr. Guyon's laboratory, CIRAD-CP). The complete system was run under Isi 205 software, developed by SHENK and WESTERHAUS (1991).

Three hundred and thirty-one delinted and crushed cotton seed samples from various places (Mali, Burkina-Faso, Togo, Senegal, Thailand, Benin and Brazil) were used for calibration.

Moreover, 82 cotton seed samples prepared under the same conditions as those used for calibration were also used to validate the method.

Near infrared absorption measurements were carried out by reflection between 1,100 and 2,500 nm for each crushed and non-conditioned sample (relative humidity not controlled). The spectrum of each sample was the mean of the 32 spectra obtained during rotation of the measuring chamber.

The «Russian» method was used to determine the oil content of each of the seed samples used to calibrate and validate the spectrophotometric method (BOURÉLY, 1982).

Calibration methods

Calibration was carried out from a file of 331 spectra, structured by the following statistical procedures.

- **Principal components analysis (PCA).** PCA is used to reduce the number of real variables, i.e. the number of wavelengths between 1,100 and 2,500 nm, to a compatible number of variables, so that the Mahalanobis distance calculations are not too laborious.

- **Generalized Mahalanobis distance (WILLIAMS and NORRIS, 1987).** For a given spectrum, this distance is its distance in absorption units compared to the mean of all the spectra (centre of gravity) for one or more given wavelengths. These distances, measured in terms of «H», should remain less than 3 (WESTERHAUS and SCHENK, 1991).

In our study, 5 samples, for which «H» was less than 3, were considered as atypical and rejected from the file.

In order to determine the most representative samples of the population to be used for calibration, statistical selection was carried out by calculating the Mahalanobis distances, so as to identify and select the spectra with the greatest number of nearest neighbours.

Mathematical processing enabled the selection of 244 samples for calibration, which were also quantified using the reference method.

Calibration enables a mathematical value to be associated with each spectrum, which itself is associated with the reference value.

Results and discussion

Determining the calibration equation

The Isi 205 software offers the possibility of choosing from several parameters that influence the effectiveness of calibration (choice of mathematical processing, regression calculation method, wavelength range used, number of cross-checks, etc.). The selected parameters that enabled us to obtain the best results were as follows:

- spectrum range, 1,100 nm - 2,500 nm,
- mathematical processing:
 - second derivative
 - step, 5 nm
 - smoothing, 5 and 1
- MPLS (modified partial least square) regression method,
- cross-checking, 4,
- number of terms in the prediction equation, 12.

The MPLS method reduces errors due to matrix calculation of Mahalanobis distances.

The spectral variations due to surface effects, which occur when the reflectance signal is used directly, can be reduced by using the first and second derivatives of Log (1/R) (COOPER, 1983; NORRIS and WILLIAMS, 1984). Using second derivatives usually gives better results (DARDENNE and BISTON, 1986).

Although strict texture conditions for product crushing are a preponderant factor (OSBORNE, 1983), using second derivatives largely smooths out the differences.

The spectrum for each sample is recorded in 5 nm steps. In order to ensure the best possible accuracy, the step should be as small as possible (fig. 2).

Successive smoothing reduces background noise.

The calibration procedure consists in choosing three quarters of the samples on which a multiple regression model is developed (MPLS). This is then applied to the remaining quarter of samples. The procedure is repeated three times, retaining a quarter of the samples each time, for validation. The deviations for the four predictions are cumulated to obtain a standard deviation for validation, the minimum of which fixes the number of factors to be introduced into the model. The final model is then recalculated on all the samples.

The maximum number of terms in the equation was fixed at 12 by the experimenter. During cross-checks, the software chooses the number of terms required to draw up

the best calibration equation. A good evaluation of the number of terms necessary is one for 10 samples.

Cross-checking further eliminated a certain number of spectra, whose deviation between the predicted values and the reference values was greater than 2% oil (value of $\pm T$ fixed by the experimenter). Ten samples with too large an $\pm H$ or $\pm T$ were rejected. The calibration results are given in table 1.

Validation

Calibration was validated by analyzing 82 independent samples.

The coefficient of correlation between the reference values and the predicted values was 0.888, i.e. a coefficient of determination of 78.9%.

Validation of this calibration showed that the correlation between the reference values and the predicted values was not bad, but insufficient for this prediction equation to be used.

Proposed method improvements

Given that the equation parameters were determined from the entire spectrum of each sample, the variation in water content may have had an effect on spectrophotometric analysis.

We suggest proceeding with fresh calibration of the oil content determination using 120 cotton seed samples from a genotype bank, conditioned at 0 or 80% relative humidity.

Cotton seed grinding should be standardized: around 20 g of seeds should be ground in a coffee grinder in six 5-second bursts, between which the material should be stirred with a spatula.

The oil content should be determined by the reference method (\pm Russian \pm method), with 8 analyses, taking the mean of the results so as to have a relative deviation of $\pm 0.25\%$ oil. The value of $\pm T$ (tolerated difference between the values obtained and the predicted values) will be chosen in accordance with the relative deviation seen in the results obtained using the reference method.

Calibration should be validated on 40 samples not used for calibration and the correlation between the NIR values and the reference values should be determined. If calibration

proves unsatisfactory, the spectral data for these 40 samples should be returned to the initial file and fresh calibration should be carried out. The calibration should then be validated with another 40 samples, etc.

It is likely that the number of samples necessary to obtain better calibration will be over 120.

Conclusion

The results of this preliminary study show that it is possible to determine the oil content of cotton seeds by NIRS.

Nevertheless, although the calibration characteristics obtained by <cross-checking> are satisfactory ($RSQ > 0.9$), validation of the method reveals too low a level of correlation between the reference values and the NIRS values for laboratory requirements.

For calibration and subsequent analyses, the method for preparing and conditioning the samples, whose physical

characteristics affect NIRS analyses, needs to be further standardized.

Our laboratory is currently working to improve this spectrophotometric method and could be using it regularly for oil content determination in varietal trials by the end of 1994.

There are a very wide range of possibilities for the use of near infrared spectrophotometry in organic component quantification. We are already planning to use it to determine the protein and gossypol content of cotton seeds.

Determinación del contenido de aceite de las semillas de algodón por espectrofotometría en el infrarrojo próximo

C. Marquié, A.-M. Tessier

Resumen

Los autores dan los resultados obtenidos en un estudio preliminar sobre la calibración de un espectrofotómetro en el infrarrojo próximo para determinar el contenido de aceite en las semillas de algodón. La validación de la calibración pone de

manifiesto una correlación insuficiente entre los valores de referencia y los obtenidos por espectrofotometría (coeficiente de determinación: 0.79). Los autores proponen modificaciones para mejorar el acondicionamiento y la medida de las muestras.

PALABRAS CLAVE: semillas de algodón, aceite, espectrofotometría de infrarrojo próximo.